Paralleled Fuzzy Time Series and Genetic Algorithm Prediction

Ankit Sirmorya, *CISE,* Preethu Thomas, *CISE,*
Shantanu Kande, *CISE,* and Sumeet Pande, *CISE*

Abstract—Paralleled Fuzzy Time Series(FTS) and Genetic algorithm(GA) prediction intends to bring these algorithms together to create a prediction model over time based events, while leveraging the MapReduce paradigm to enable parallel execution. Fuzzy Time Series is a algorithm that allows prediction by developing a model based on historical time based events. Genetic Algorithm emulates natural selection, such that we can improve the forecasts of a given prediction model. This process is run till the prediction converges sufficiently with the training data.

Index Terms—Fuzzy Time Series, Genetic algorithm, MapReduce

1 Introduction

Accurately forecasting results for a given use case is always a challenge. It depends on multiple variables, there are exceptions which are not the norm that occur in everyday life and all related data for the given event may not be available.

An algorithm can make "as close to the truth" prediction, if it has sufficient historical data and it makes sense and relates the available data to the pattern of occurrences of that event. Weather forecasting, Election predictions, Stock market analysis and recommendations are few of the numerous examples that we see in everyday life, where prediction models are effectively used.

In this project, we intend to use Fuzzy Time Series (FTS) [1] and Genetic Algorithm (GA) [5] prediction models and translate them into MapReduce paradigm for reliable, accurate and efficient forecasting of time based events.

The novelty of this implementation is to provide an end-to-end solution by bringing together FTS, GA and MapReduce. FTS and GA being used in conjunction should improve prediction results, while MapReduce should improve running time as well as increase the amount of data ingested.

2 Architecture

The architecture is divided into three modules data ingestion, prediction and data visualization. This is as shown in *Fig 1.*

3 Use Cases

The approach introduced in this paper will enhance the prediction of any time base events such as accidents, election campaigns, sales and so forth.

4 Algorithm

The algorithm for this project:

Step 1: Read the training data which in this case will be time based events.

Step 2: Find the maximum and the minimum values of the training data and compute the Universe of Discourse.

Step 3: Create the first generation of Genetic Algorithm which will be a set of randomized individuals. Each of the individuals will be within the Universe of Discourse.

Step 4: Create the division of the generation data which can be used in the MapReduce model.

Step 5: The mapper task will apply Fuzzy Time Series Algorithm using the training data and GA population. This will make prediction for each individual and fitness value for each

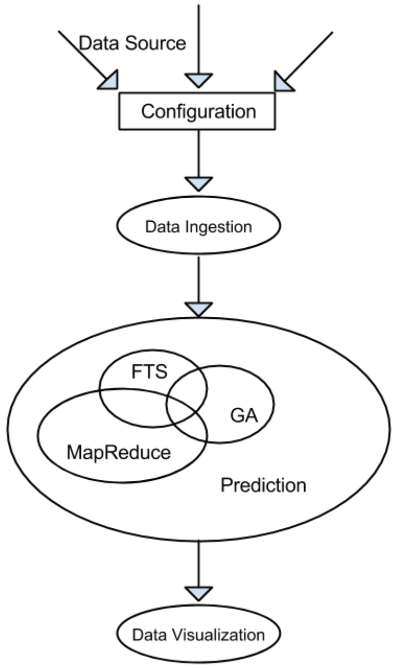

Fig. 1. Architecture of the implementation

of them will be calculated.

Step 6: The reducer task will perform the selection, cross-over and mutation operations of GA based on the fitness value calculated by the mapper task. Thus, a new generation of individuals will be created.

Step 7: Repeat Steps 4 through 6 until the convergence point is not reached.

5 Modules

There are three major modules namely data ingestion, prediction and data visualization.

Data ingestion will allow a configurable way to parse the input data. On existence of necessarily columns for the prediction, the module will execute. Initially, data will be read from MSExcel format.

Prediction module will include implementation of FTS and GA algorithms in the MapReduce paradigm.

Data visualization module will use the prediction results in various output formats to easily understand and visualize the results.

6 Flowchart

Fig. 2. is the flowchart for this implementation

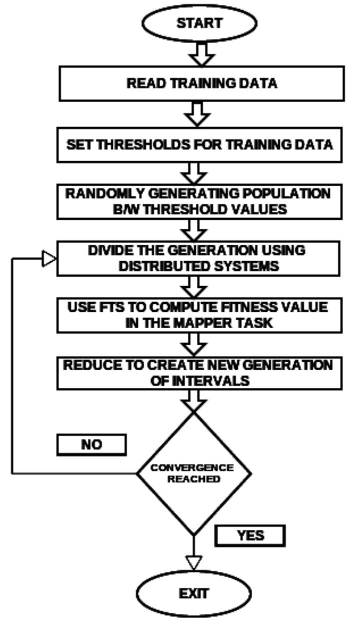

Fig. 2. Flowchart

7 Potential Result

Some of the potential results from this implementation are more accurate predictions, quicker results and comparison to related work.

As FTS and GA are used in an amalgamated way, the expectation is that the prediction model will be more refined in comparison to how FTS performs individually. Furthermore, the use of MapReduce will help speedup the process and help glean quicker running time.

The quantification of "how correct are the predictions?" would be evident by using historical data as training, test data and comparing how the predicted value is different from the test value. Finally, we will compare our results with results of related work which could showcase the efficiency of our model in running time and over large data sets.

8 Related Work

Fuzzy Time Series (FTS) [1] is a prediction model that has proved its ability to forecast in numerous use cases, such as tourism [2], enrollments [3], temperature [4] and many more.

Holland's Genetic algorithm [5] has also been successfully been used in multiple prediction examples [6], [7], [11] and also in conjunction with FTS [8]. Genetic algorithm has also been run as MapReduce to solve the OneMax problem [10].

9 Novelty of the Idea

Though there have been previous researches done in the field of parallel evolutionary algorithm [12], in this paper we have tried to overcome the time incurred by GA-FTS prediction model.

Also, with the advent of MapReduce [9] paradigm, it is much easier to abstract parallel system operations; while paving the way to concentrate on the problem at hand. However, a restrictive programming model like MapReduce, requires the translation of the algorithm in the form of map/reduce functions. This is still a major challenge and needs to be done with much consideration.

10 Possibilities to Consider

During the execution of the project, the following will still be in consideration:

Can we dynamically retrieve data? Initially, we will be manually downloading/creating the data.

Can we use some other parallel framework? Possible candidates could be Spark [13], Storm [14], Giraph [15] or Hama [16]

Can we employ different MapReduce models of GA? Initially, only selection, crossover and mutation stages will be in reducer phase. We need to evaluate different mechanisms to improve the reduce phase.

Can we dynamically set thresholds for FTS/GA and number of individuals in GA? We will be manually tweaking these threshold values and visualizing the effects. We will try to have a dynamic way to sample training data and running multiple instance of the algorithm in parallel to decide appropriate threshold values.

References

[1] Q. Song, B. Chissom, *Fuzzy time series and its models*, Fuzzy Sets and Systems, Vol 54, 1993, p. 269-277.

[2] C.H, Wang and L.C. Hsu, *Constructing and applying an improved fuzzy time series model: Taking the tourism industry for example*, Expert Systems with Applications, Vol 34, 2008, p. 2732-2738.

[3] S.M. Chen, *Forecasting enrollments based on high-order fuzzy time series*, Cybernetics and Systems: An International Journal, Vol 33, 2002, p. 1-16. 781-789.

[4] S.M. Chen and J.R. Hwang, *Temperature prediction using fuzzy time series*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol 30, Issue 2, 2000, p. 263-275.

[5] John H. Holland, *Adaptation in Natural and Artificial Systems*, 1975, University of Michigan Press, Ann Arbor.

[6] K.S. Shin and Y.J. Lee, *A genetic algorithm application in bankruptcy prediction modeling*, Expert Systems with Applications, 2002, 321-328

[7] A. Azadeh, S.F. Ghaderi, S. Tarverdian and M. Saberi. *Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption*, Applied Mathematics and Computation, 2007, 186:1731-41

[8] S.M. Chen and N.Y. Chung, *Forecasting enrollments using high-order fuzzy time series and genetic algorithms*, Internat. J. Intell. Syst. 21 ,2006, 485-501.

[9] Jeffrey Dean and Sanjay Ghemawat, *MapReduce: Simplified Data Processing on Large Clusters*,OSDI'04: Sixth Symposium on Operating System Design and Implementation, 107-113.

[10] A. Verma, X. Llora, D.E. Goldberg, and R.H.Campbell, *Scaling genetic algorithms using mapreduce*, Intelligent Systems Design and Applications, 2009, 13-18.

[11] Deepak Singh and Ankit Sirmorya, *Article: Solving Real Optimization Problem Using Genetic Algorithm with Employed Bee (GAEB)*, International Journal of Computer Applications, 2012, 1-5.

[12] E. Cant-Paz, *A survey of parallel genetic algorithms*, Calculateurs paralleles, reseaux et systems repartis, 1998, 141-171.

[13] Apache Spark - *https://spark.apache.org/*

[14] Apache Storm - *https://storm.incubator.apache.org/*

[15] Apache Giraph - *http://giraph.apache.org/*

[16] Apache Hama - *https://hama.apache.org/*